EMBEDDED INTER-SUBJECT VARIABILITY IN ADVERSARIAL LEARNING FOR INERTIAL SENSOR-BASED HUMAN ACTIVITY RECOGNITION

Anonymous

Anonymous

ABSTRACT

This paper addresses the problem of Human Activity Recognition (HAR) using data from wearable inertial sensors. An important challenge in HAR is the model's generalization capabilities to new unseen individuals due to inter-subject variability, i.e., the same activity is performed differently by different individuals. To address this problem, we propose a novel deep adversarial framework that integrates the concept of inter-subject variability in the adversarial task, thereby encouraging subject-invariant feature representations and enhancing the classification performance in the HAR problem. Our approach outperforms previous methods in three well-established HAR datasets using a leave-one-subject-out (LOSO) cross-validation. Further results indicate that our proposed adversarial task effectively reduces inter-subject variability among different users in the feature space, and it outperforms adversarial tasks from previous works when integrated into our framework. Code:`https://anonymous.4open.science/r/MLSP2025-7D97`

***Index Terms*—** Human Activity Recognition, Inertial Wearable Sensors, Inter-Subject Variability, Deep Adversarial Learning

1. APPENDIX A

This section introduces the architecture details for the feature extractor F [1], the activity classifier C (Table 1), the discriminator D (Table 2), and the reconstructor R (Table 3).

Table 1: Parameters for our activity classifier C. We use ReLU activation function after Linear 1 and 2. We use Softmax function after Linear 3.

Layer	Weights [in, out]	Activation
Linear 1	[64, 512]	ReLU
Linear 2	[512, 256]	ReLU
Linear 3	[256, $\lvert Y \rvert$]	Softmax

Table 2: Parameters for our discriminator D. After each convolution, there is a batch normalization+ReLU. **Legend**: OSh (Output Shape), Ker (Kernel), Str (Stride), Dr (Dropout).

Layer	**OSh**	**Ker**	**Str**	**Dr**
ConvBlock 1	[32, 60]	[1,9]	[1,2]	-
ConvBlock 2	[64, 28]	[1,5]	[1,2]	-
ConvBlock 3	[128, 13]	[1,3]	[1,2]	-
Dropout	[128, 13]	-	-	0.3
ConvBlock 4	[256, 6]	[1,3]	[1,2]	-
Dropout	[256, 6]	-	-	0.3
Flatten	[1536]	-	-	-
Linear	[256]	-	-	-
Dropout	[256]	-	-	0.2
Linear	[64]	-	-	
Linear	[1]	-	-	

Table 3: Parameters for our reconstructor. After each convolution, there is a batch normalization+LEakyReLU with a negative slope of 0.01. The variables c and w denote the number of sensor channels and the window size respectively. **Legend**: OSh (Output Shape), Ker (Kernel), Str (Stride), Pad (Padding), Dil (Dilation), OPa (Output Padding).

Layer	**OSh**	**Ker**	**Str**	**Pad**	**Dil**	**OPa**
Linear	[1, 1, 128]	-	-	-	-	-
ConvTranspose2d	[128, 1, 32]	[1,3]	[1,1]	[0,1]	16	[0,1]
Conv2d	[128, 1, 32]	[1,3]	[1,1]	[0,1]	-	-
ConvTranspose2d	[64, 1, 64]	[1,3]	[1,1]	[0,1]	16	[0,2]
Conv2d	[64, 1, 64]	[1,3]	[1,1]	[0,1]	-	-
ConvTranspose2d	[c, 1, 128]	[1,5]	[1,2]	[0,2]	1	[0,1]
ConvTranspose2d	[c, 1, 256]	[1,5]	[1,2]	[0,2]	1	[0,1]
ConvTranspose2d	[c, 1, w]	[1,5]	[1,2]	[0,2]	1	[0,1]

Anonymous.

979-8-3503-2411-2/25/$31.00 ©2025 IEEE

2. REFERENCES

[1] Francisco M. Calatrava-Nicolás and Oscar Martinez Mozos, "Light residual network for human activity recognition using wearable sensor data," *IEEE Sensors Letters*, vol. 7, no. 10, pp. 1–4, 2023.